

### **Listing of the Claims**

Claims 1-30 were pending.

Please amend claims 2-8 and 10-18.

Kindly cancel claims 1 and 19-30 without prejudice.

Please add claims 72 and 73.

Accordingly claims 2-18 and 72-73 remain pending.

1. (Canceled)
2. (Currently amended) The A method ~~as recited in~~ of claim 9 +, wherein the frame of content comprises a frame of video content.
3. (Currently amended) The A method ~~as recited in~~ of claim 9 +, wherein the frame of content comprises a frame of audio content.
4. (Currently amended) The A method ~~as recited in~~ of claim 9 +, wherein the frame of content comprises a frame of both video and audio content.
5. (Currently amended) The A method ~~as recited in~~ of claim 9 +, further comprising repeating the automatically detecting in the event tracking of a verified face is lost.

6. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1,~~  
wherein the method further comprises receiving the frame of content ~~comprises~~  
~~receiving a frame of video content~~ from a video capture device local to a system  
implementing the method.

7. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1,~~  
wherein the method further comprises receiving the frame of content ~~comprises~~  
~~receiving the frame of content~~ from a computer readable medium accessible to a  
system implementing the method.

8. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1,~~  
wherein automatically detecting the candidate area further comprises:

~~detecting whether there is motion in the frame and, if there is motion in the~~  
~~frame, then performing motion-based initialization to identify one or more~~  
~~candidate areas;~~

detecting whether there is audio in the frame, and if there is audio in the  
frame, then performing audio-based initialization to identify one or more  
candidate areas; and

using, if there is neither motion nor audio in the frame, a fast face detector  
to identify one or more candidate areas.

9. (Previously Presented) A method comprising:  
receiving a frame of content;

automatically detecting a candidate area for a new face region in the frame, wherein detecting the candidate area comprises:

- determining whether there is motion at a plurality of pixels on a plurality of lines across the frame;

- generating a sum of frame differences for each possible segment of each of the plurality of lines;

- selecting, for each of the plurality of lines, the segment having the largest sum;

- identifying a smoothest region of the selected segments;

- checking whether the smoothest region resembles a human upper body; and

- extracting, as the candidate area, a portion of the smoothest region that resembles a human head;

- using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

- indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

- using a plurality of cues to track each verified face in the content from frame to frame.

10. (Currently amended) The A method as recited in of claim 9,  
wherein determining whether there is motion comprises:

determining, for each of the plurality of pixels, whether a difference between an intensity value of the pixel in the frame and an intensity value of a corresponding pixel in one or more other frames exceeds a threshold value.

11. (Currently amended) The A method as recited in of claim 9 ~~1~~, wherein the one or more hierarchical verification levels include a coarse level and a fine level, wherein the coarse level can verify whether the human face is in the candidate area faster but with less accuracy than the fine level.

12. (Currently amended) The A method ~~as recited in~~ of claim 9 4,  
wherein using one or more hierarchical verification levels comprises, as one of the  
levels of verification:

generating a color histogram of the candidate area:

generating an estimated color histogram of the candidate area based on previous frames;

determining a similarity value between the color histogram and the estimated color histogram; and

verifying that the candidate area includes a face if the similarity value is greater than a threshold value.

13. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1~~, wherein indicating that the candidate area includes the face comprises recording the candidate area in a tracking list.

14. (Currently amended) ~~The A method as recited in~~ of claim 13, wherein recording the candidate area in the tracking list comprises accessing a record corresponding to the candidate area and resetting a time since last verification of the candidate.

15. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1~~, wherein the one or more hierarchical verification levels include a first level and a second level, and wherein using the one or more hierarchical verification levels to verify whether the human face is in the candidate area comprises:

checking whether, using the first level verification, the human face is verified as in the candidate area; and

using the second level verification only if the checking indicates that the human face is not verified as in the candidate area by the first level verification.

16. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~1~~, wherein using one or more hierarchical verification levels comprises:

using a first verification process to determine whether the human head is in the candidate area; and

if the first verification process verifies that the human head is in the candidate area, then indicating the area includes a face, and otherwise using a second verification process to determine whether the human head is in the area.

17. (Currently amended) ~~The A method as recited in~~ of claim 16, wherein the first verification process is faster but less accurate than the second verification process.

18. (Currently amended) ~~The A method as recited in~~ of claim 9 ~~+~~, wherein the plurality of cues include foreground color, background color, edge intensity, motion, and audio.

19-71. (Canceled)

72. (New) A computer-readable storage medium comprising computer-program instructions that when executed by a processor perform acts of:

receiving a frame of content;

automatically detecting a candidate area for a new face region in the frame,

wherein detecting the candidate area comprises:

determining whether there is motion at a plurality of pixels on a plurality of lines across the frame;

generating a sum of frame differences for each possible segment of each of the plurality of lines;

selecting, for each of the plurality of lines, the segment having the largest sum;

identifying a smoothest region of the selected segments;

checking whether the smoothest region resembles a human upper body; and

extracting, as the candidate area, a portion of the smoothest region that resembles a human head;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame.

72. (New) A computing device comprising:

a processor; and

a memory coupled to the processor, the memory comprising computer-program instructions that when executed by the processor perform acts of:

receiving a frame of content;

automatically detecting a candidate area for a new face region in the frame, wherein detecting the candidate area comprises:

determining whether there is motion at a plurality of pixels on a plurality of lines across the frame;

generating a sum of frame differences for each possible segment of each of the plurality of lines;

selecting, for each of the plurality of lines, the segment having the largest sum;

identifying a smoothest region of the selected segments;

checking whether the smoothest region resembles a human upper body; and

extracting, as the candidate area, a portion of the smoothest region that resembles a human head;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame.